

A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions

CHAD J. CREIGHTON,^{1,2} ANKUR K. NAGARAJA,^{3,4} SAMIR M. HANASH,⁵ MARTIN M. MATZUK,^{3,4,6} and PREETHI H. GUNARATNE^{3,7}

¹The Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA

²Department of Medicine, Baylor College of Medicine, Houston, Texas 77030, USA

³Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA

⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

⁵Molecular Diagnostics, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

⁶Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

⁷Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA

ABSTRACT

MicroRNAs are short (~22 nucleotides) noncoding RNAs that regulate the stability and translation of mRNA targets. A number of computational algorithms have been developed to help predict which microRNAs are likely to regulate which genes. Gene expression profiling of biological systems where microRNAs might be active can yield hundreds of differentially expressed genes. The commonly used public microRNA target prediction databases facilitate gene-by-gene searches. However, integration of microRNA–mRNA target predictions with gene expression data on a large scale using these databases is currently cumbersome and time consuming for many researchers. We have developed a desktop software application which, for a given target prediction database, retrieves all microRNA:mRNA functional pairs represented by an experimentally derived set of genes. Furthermore, for each microRNA, the software computes an enrichment statistic for overrepresentation of predicted targets within the gene set, which could help to implicate roles for specific microRNAs and microRNA-regulated genes in the system under study. Currently, the software supports searching of results from PicTar, TargetScan, and miRanda algorithms. In addition, the software can accept any user-defined set of gene-to-class associations for searching, which can include the results of other target prediction algorithms, as well as gene annotation or gene-to-pathway associations. A search (using our software) of genes transcriptionally regulated *in vitro* by estrogen in breast cancer uncovered numerous targeting associations for specific microRNAs—above what could be observed in randomly generated gene lists—suggesting a role for microRNAs in mediating the estrogen response. The software and Excel VBA source code are freely available at <http://sigterms.sourceforge.net>.

Keywords: microRNA; gene expression profiling; PicTar; TargetScan; miRanda

INTRODUCTION

MicroRNAs are small ~22-nucleotide RNAs that constitute an extensive class of noncoding RNAs that direct the translational repression and/or cleavage of complementary target mRNAs (Du and Zamore 2005). MicroRNAs govern

broad gene regulatory networks and are essential for normal mammalian development, including roles in cell growth, differentiation, and apoptosis (Alvarez-Garcia and Miska 2005). Alterations in microRNA expression have also been observed in a variety of human cancers (Lu et al. 2005; Volinia et al. 2006; Ozen et al. 2008).

The formation of a double-stranded RNA duplex through the binding of microRNA to mRNA in the RNA induced silencing complex (RISC) either triggers the degradation of the mRNA transcript or the inhibition of protein translation. Because experimental identification of microRNA targets is difficult, a number of computational algorithms have been developed to predict the targeting of

Reprint requests to: Chad J. Creighton, Dan L. Duncan Cancer Center Division of Biostatistics, Baylor College of Medicine, One Baylor Plaza M.S. 305, Houston, TX 77030, USA; e-mail: creightc@bcm.edu; fax: (713) 798-2716.

Abbreviations: miRNA, microRNA; E2, 17 β -estradiol.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1188208>.

a given mRNA by a specific microRNA. The first step of prediction involves identifying potential microRNA binding sites in the mRNA according to specific base-pairing rules, and the second step involves implementing cross-species conservation requirements, with different algorithms using slightly different criteria for each step (Sethupathy et al. 2006). Some of the more popular prediction algorithms include PicTar (Krek et al. 2005), TargetScan (Lewis et al. 2005), and miRanda (John et al. 2004). Each algorithm has a sizable rate of both false positive and false negative predictions (Rajewsky 2006), and thus more than one algorithm may be necessary to make predictions about any particular gene or microRNA.

Gene expression profiling at either the protein or mRNA level can often reveal hundreds of gene expression changes for a biological system of interest (Creighton et al. 2006). Up to 30% of the genes in mammalian genomes have been predicted to be regulated by microRNAs (Lewis et al. 2003; John et al. 2004). Therefore, microRNA regulation is highly likely to underlie many of the differential expression patterns being observed. Published gene targeting prediction databases are often made available via a web interface, where the user can look up predicted microRNA:mRNA functional pairs for a specific microRNA or gene of interest. In cases where the number of genes of interest (e.g., a set of genes arising from a set of expression profiling experiments) is on the order of tens or even hundreds, a gene-by-gene approach to looking up microRNA:mRNA pairs becomes impractical. Furthermore, when considering a sizable number of genes that may be coordinately regulated, one may desire to evaluate the genes as a group where multiple mRNAs have binding sites for one or specific microRNAs or families of microRNAs; this may be done using classical enrichment statistics testing overrepresentation of the microRNA target predictions within the selected set of genes.

We have developed a desktop software application for use with Microsoft Excel that, for a given target prediction database, retrieves all microRNA:mRNA functional pairs represented by an input set of genes. Furthermore, for each microRNA, the software computes an enrichment statistic for overrepresentation of predicted targets within the gene set. The entire set of microRNA:mRNA predictions from PicTar, TargetScan, and miRanda databases have been parsed into the required format for use with the software and have been made available. Our software application is flexible in that predictions from other algorithms can also be integrated, as long as these predictions are precompiled into a file in the required format. In addition to gene-to-microRNA associations, the software can accept any user-defined set of gene-to-class associations for searching, which can include gene annotation or gene-to-pathway associations.

The software, source code, and auxiliary files (including tutorial) are freely available at <http://sigterms.sourceforge.net>.

Similar to a number of other successful tools for molecular profile analysis (Tusher et al. 2001; Tibshirani et al. 2002; Xu et al. 2008)—including “SAM” and “BRB-ArrayTools”—our software package utilizes Microsoft Excel as the front end. Most scientists are familiar with Excel, allowing even researchers who do not have programming or scripting skills to easily use our software. We demonstrate the software here, using sets of genes regulated by estrogen from a previously published study (Creighton et al. 2006).

RESULTS

Software for linking gene lists to a set of microRNA target predictions

Our software is comprised of a set of Excel macros for use with Microsoft Excel. Given a user-specified list of genes (e.g., a set of genes significantly up-regulated in a particular expression profiling experiment) and a set of gene-to-microRNA pair associations (e.g., the entire set of microRNA:mRNA target predictions from TargetScan), the software will retrieve all gene-to-microRNA pairs that involve the genes in the user-specified list. The gene-to-microRNA pairs are read into the software from an Excel workbook in a specific format, which we refer to here as the “Annotation” workbook. The Annotation workbook contains the gene annotation information (which may include microRNA targeting predictions) for the entire population of genes under study (e.g., the set of genes represented on the expression profiling platform).

The Annotation workbook includes a spreadsheet with genes listed in the rows (one gene per row) and microRNA associations (or other types of gene class associations) listed in the columns adjacent to the corresponding gene. The same gene may be listed in multiple rows, and so the maximum number of microRNA associations for a given gene is not limited to the maximum number of columns in an Excel spreadsheet. The maximum number of rows in a spreadsheet (65,536 for 32-bit Excel, ~1 million for Excel Vista) greatly exceeds the number of unique named genes for a given organism. The Annotation workbook format allows for a theoretical limit of ~16 million gene-to-microRNA associations in 32-bit Excel (i.e., slightly less than the number of cells in a spreadsheet). For the purposes of carrying out enrichment tests (described below), the Annotation workbook includes another spreadsheet listing the total number of genes targeted for each microRNA.

For the purposes of retrieving microRNA:mRNA pairs, the Annotation workbook could represent a particular target prediction database (e.g., TargetScan). We have compiled the entire set of predictions for each of the widely used PicTar, TargetScan, and miRanda databases into the Annotation workbook format (for both human and mouse); these files are available at the software website,

<http://sigterms.sourceforge.net>. Annotation workbooks for additional target prediction databases may also be made available through the website as the predictions become available and as general user interest warrants. Additionally, users have the freedom to construct their own Annotation workbooks for their gene class associations of interest (if these are not represented in the precompiled workbooks that we provide), though the typical user will not have a need to do this and can simply download the Annotation workbook of interest from the software website.

Testing microRNAs for enrichment of predicted gene targets within gene lists

In addition to retrieving the set of microRNA:mRNA functional pairs involving each individual gene within a given set, one may wish to evaluate the genes as a group for targeting by a specific microRNA. For a given microRNA, the question that can be asked is whether the microRNA targets a higher number of genes within the selected gene set than would be expected in a random gene set. Statistics are needed to factor in the absolute numbers of genes involved, in addition to the proportional differences in gene class representation between the selected set and the entire gene population. Testing gene sets for enrichment of specific microRNA associations is entirely analogous to what has previously been done in testing gene sets for gene annotation (e.g., the “GO term” [Gene Ontology]) enrichment (Creighton et al. 2003; Doniger et al. 2003). As is common practice for testing GO term enrichment, our software computes the extent of gene target enrichment for each retrieved microRNA, using the classical one-sided Fisher’s exact test.

For precise calculation of the enrichment statistic, the user specifies the total population of genes from which the selected gene set was obtained (which, in the case of gene expression profiling, could be taken to be the total number of genes represented on an array platform). As a rule, any genes not in the total population should not be represented in the Annotation workbook. For commonly used array platforms for mouse and human (including Affymetrix and Illumina), precompiled Annotation workbooks specific to the given platform are available from the software website (which include the number of unique genes represented on the given array).

The one-sided Fisher’s exact *P*-value gives the probability of enrichment for that particular microRNA. However, where several hundred microRNAs are simultaneously considered for enrichment, the issue of multiple testing needs to be taken into account. One way to address this is to use methods such as from Storey and Tibshirani (2003) to estimate the false discovery rate (FDR) (i.e., multiply the total number of microRNAs tested by a given nominal *P*-value, then divide by the total number of microRNAs having *P*-values less than the given). The Storey method,

however, assumes that microRNAs are generally independent from each other in terms of the genes that they can target.

Another way to address multiple testing is to perform Monte Carlo simulations, where, in each simulation, a random set of genes of the same number as the genes in the actual set are generated and tested for enriched microRNA associations. One can then examine the distribution of nominal enrichment *P*-values generated from each simulation test to estimate the number of microRNAs that may have received a low *P*-value by chance alone. Our software includes a feature for Monte Carlo simulation testing, where the user inputs the size of the gene set, the list of genes in the total population, and the number of simulations to carry out. The nominal *P*-values generated for each test are then outputted to one or more Excel spreadsheets for comparison with the *P*-values arising from the actual gene set.

Linking gene lists to other types of gene class associations

In addition to gene-to-microRNA associations, the software can accept any user-defined set of gene-to-class associations for retrieval and enrichment testing. Other types of gene information, such as GO annotations, can be formatted as Annotation workbooks for input into the software. The software website currently provides preformatted Annotation workbooks for searching GO terms (Gene_Ontology_Consortium 2004), gene sets from the Molecular Signature Database (MSigDb) (Subramanian et al. 2005), and a set of “oncogenic” gene expression signatures recently collected using public profiling data sets (Creighton 2008). Users are free to define additional Annotation workbooks of their own as well. The modular nature of the Annotation workbooks gives the user substantial flexibility in choosing or defining the types of gene-to-class associations to be searched.

Genes regulated by estrogen are enriched for target predictions for specific microRNAs

Estrogenic hormones are key regulators of growth and differentiation and function in a wide array of target tissues. Estrogens also play a central role in breast cancer pathogenesis. Previously, gene expression profiles were generated from three estrogen receptor (ER)-positive breast cancer cell lines stimulated by 17 β -estradiol (E2) *in vitro* over a time course (Creighton et al. 2006). Here, we focused on the published sets of gene transcripts induced or repressed early by E2 (within 4 h), and with sustained induction or repression through 24 h (the previously designated gene clusters “B” and “F,” respectively). Our goals were to determine what microRNAs might be associated with these E2-regulated genes and whether

predicted target sites for specific microRNAs might be significantly enriched within the 3' UTRs of E2-regulated gene sets as a group.

We used our software to find all predicted microRNA:mRNA functional pairs involving any of the genes in the set induced by E2 and in the set repressed by E2. We searched the set of E2-induced genes separately from the E2-repressed genes for predicted target sites for microRNAs in each of the PicTar, TargetScan, and miRanda databases. For the 623 unique E2-induced genes, we found 2087, 2181, and 21,853 microRNA:mRNA pairs for PicTar, TargetScan, and miRanda, respectively. For the 268 E2-repressed genes, we found 1207, 1097, and 9647 microRNA:mRNA pairs for PicTar, TargetScan, and miRanda, respectively. The miRanda algorithm uses more relaxed criteria (presumably resulting in a higher false positive rate but a lower false negative rate) as compared to PicTar or TargetScan (Sethupathy et al. 2006), which accounts for the greater number of retrieved predictions for miRanda; in addition, more microRNAs were represented in miRanda over PicTar or TargetScan (395 microRNAs compared with 164 microRNAs and 162 microRNA families, respectively).

Using the software, we evaluated each microRNA (included in the given target prediction database) for overrepresentation within the E2-regulated gene set, as compared to the entire set of microRNA:mRNA pairs for that microRNA involving all genes represented on the entire array. In each of the three prediction databases, we found numerous microRNAs that were overrepresented ($P < 0.05$, one-sided Fisher's exact test) in either the E2-induced or the E2-repressed gene sets (Fig. 1). As multiple microRNAs were each considered for enrichment, we performed simulation testing with our software to determine how many nominally significant microRNAs might occur within a random set of genes. The number of enriched microRNA associations (nominal) $P < 0.05$ in the E2 gene sets greatly

exceeded the expected number of nominally significant associations from a random gene set (Fig. 1). For each of the six gene-set/prediction-database pairs (induced and repressed versus PicTar, TargetScan, and miRanda), $<5\%$ of the 100 random gene sets had a number of nominally significant associations exceeding the number obtained using the actual gene set.

For each E2-regulated gene set (induced or repressed) and for each microRNA target prediction database, the top enriched microRNA targeting associations are listed (Fig. 2). As indicated in Figure 2, many microRNAs were significantly enriched in both the PicTar and TargetScan data sets, while many of the microRNAs significant for miRanda were not significant in either of the other two databases. The similarities in overall sets of predicted target sites between PicTar and TargetScan and the differences in the results of these algorithms from the results of miRanda have been noted previously (Rajewsky 2006).

DISCUSSION

In this article, we describe a software application that is designed for rapid large-scale identification of microRNA targeting associations in relation to entire sets of genes. The software application includes enrichment tests to determine whether a particular microRNA may be predicted to target more genes in the selected set than expected in a random set. As a demonstration of our software, we found that genes transcriptionally regulated in vitro by estrogen in breast cancer uncovered numerous targeting associations for a number of specific microRNAs. This suggests a role for microRNAs in the estrogen response, potentially through direct regulation of the mRNAs found to be differentially expressed as a result of estrogen treatment. It is conceivable that some of these microRNAs that were discovered to be significant by our enrichment analysis may

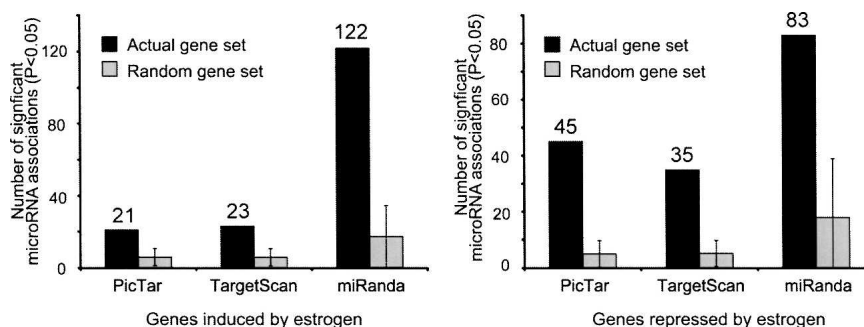
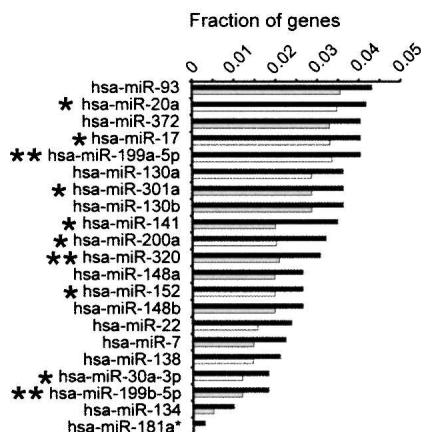
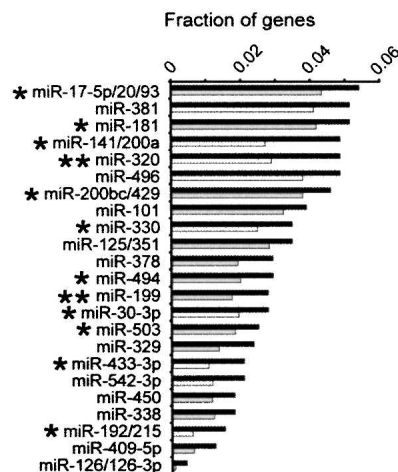
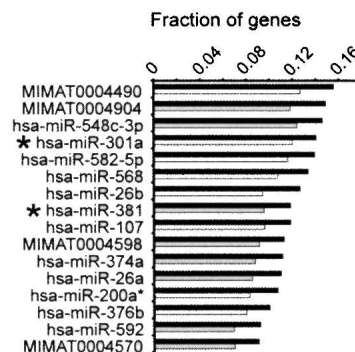


FIGURE 1. Estrogen-regulated genes are enriched in target predictions for specific microRNAs. For each microRNA:mRNA target prediction database considered (PicTar, TargetScan, and miRanda), the numbers of microRNAs that were overrepresented ($P < 0.05$, one-sided Fisher's exact test) in the given E2-regulated gene set (induced, *left* panel, or repressed, *right* panel) are shown. The expected numbers of enriched microRNA associations for a randomly generated gene set (of the same size as the actual gene set), are also given (average based on 100 simulations, error bars denote ± 1 SD).

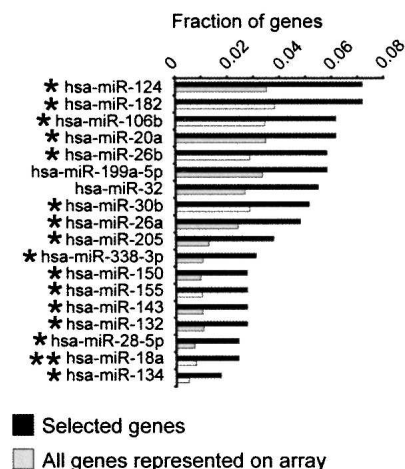
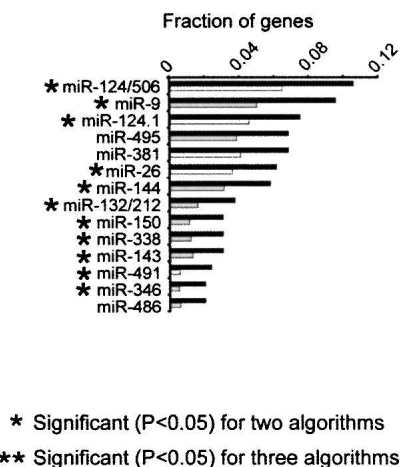
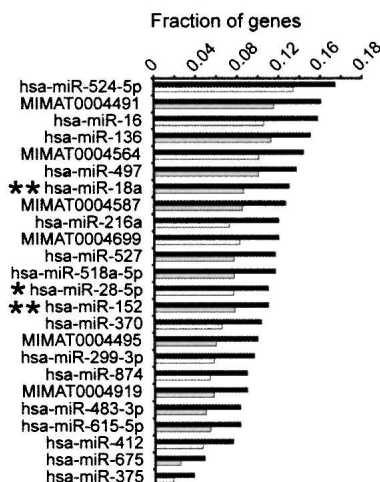
truly be regulated by estrogen. One recent study profiling zebrafish tissues found numerous microRNAs that were regulated by estrogen in that system (Cohen et al. 2008), and thus estrogen regulation of microRNAs may well be a factor in the enrichment patterns that we observe in breast cancer. In principle, however, a microRNA need not itself be differentially expressed for it to play a role in regulating mRNA and protein levels. It remains to be seen how well a pattern of enrichment of a given set of genes for a microRNA targeting association correlates with the microRNA itself also being differentially regulated.

Many microRNAs had significant associations for the estrogen-regulated

Genes induced by estrogen

PicTar ($P < 0.05$, FDR = 0.5)TargetScan ($P < 0.05$, FDR = 0.26)miRanda ($P < 0.001$, FDR = 0.02)

Genes repressed by estrogen

PicTar ($P < 0.01$, FDR = 0.06)TargetScan ($P < 0.01$, FDR = 0.07)miRanda ($P < 0.01$, FDR = 0.15)

■ Selected genes

□ All genes represented on array

* Significant ($P < 0.05$) for two algorithms** Significant ($P < 0.05$) for three algorithms

FIGURE 2. Top enriched microRNA targeting associations for each E2-regulated gene set (induced or repressed), within each of the three microRNA target prediction databases considered (PicTar, TargetScan, and miRanda). For each microRNA listed, the fraction of microRNA:mRNA pairs represented by the given gene set (black, E2-induced or E2-repressed) are compared with the fraction of microRNA:mRNA pairs represented on the entire set of genes (12,724 uniquely named) represented on the profiling array platform used in the E2 study (gray). (FDR) false discovery rate. (For miRanda, some microRNAs did not have a common name in miRBase version 11, and so accession numbers are listed for these instead.)

gene sets in two or three prediction databases. Each of the three microRNA target prediction databases that we analyzed yielded significant target associations, suggesting that each algorithm has its merits, and the use of multiple algorithms could strengthen the analysis. Each microRNA target prediction algorithm is expected to have a sizable rate of both false positive and false negative predictions (Rajewsky 2006), and by nature, defining microRNA targets is a balance between minimizing false positive and false negative rates. For instance, using more stringent criteria would likely yield fewer false positives (predicted micro-

RNA:mRNA pairs that do not validate)—i.e., higher specificity, but greater false negatives (true microRNA:mRNA pairs that were missed)—i.e., lower sensitivity.

One potential method to decrease the incidence of false positive predictions and to narrow down the list of putative microRNA targets would be to compare these *in silico* target predictions to the genes that are differentially expressed in the biological system of interest. Using the results of both microRNA and gene expression profiling, we might expect to refine predicted microRNA–mRNA associations through the identification of anti-correlated

pairs; based on what we know about microRNA function, we expect that up-regulation of a specific microRNA will lead to lower expression of its mRNA targets, and down-regulation of a specific microRNA will lead to higher levels of its target genes. Our software can aid this type of analysis approach by first allowing the user to retrieve all microRNA:mRNA pairs for the set of genes, from which (by use of a simple table join) the user can then quickly filter for those pairs involving the select set of microRNAs of interest.

MATERIALS AND METHODS

Software development

The software was developed as a set of Excel macros, using Excel Visual Basic for Applications (VBA) (~1000 lines of VBA code). To use the software, the user runs the appropriate macro, while having both the selected gene set and the entire set of gene-to-microRNA associations to be searched open as worksheets in Excel.

Target prediction data sets

The entire set of microRNA:mRNA target predictions for both human and mouse were downloaded from public websites (PicTar algorithm: <http://pictar.bio.nyu.edu>, predictions from Krek et al. 2005; TargetScan algorithm, release 4.1: <http://www.targetscan.org>; Lewis et al. 2005; miRanda algorithm from <http://www.microrna.org>; Betel et al. 2008). Each prediction data set was then parsed into an Annotation workbook (described above) for use with the software, and these workbooks are available from the software website. For searching the estrogen-regulated gene sets for targeting microRNAs, the “conserved” TargetScan predictions were used.

Analysis of estrogen-regulated gene sets

Estrogen-regulated gene sets (generated using the Affymetrix platform) were obtained from Creighton et al. (2006). The Affymetrix probe identifiers were mapped to Entrez gene identifiers, using version na.21 of the U133A annotation. The set of estrogen-induced genes was searched separately from the set of estrogen-repressed genes in each of the three target prediction databases considered (six searches in all), using our software as described above. The 12,724 uniquely named genes represented on the U133A platform were used as the reference gene population for computing the one-sided Fisher’s exact tests for enrichment of a set of targets for a particular microRNA within the set of genes. To account for multiple testing of microRNAs, Monte Carlo simulation testing using 100 randomly generated gene sets was carried out for each estrogen response group and target prediction database. For a given gene set and a given target prediction database, the number of microRNAs having a nominal significant P -value ($P < 0.05$) for target enrichment was computed for each of the 100 random tests (Fig. 1). To calculate FDR (Fig. 2), the average number of microRNA associations less than or equal to the given nominal P -value for the 100 random tests was used.

SUPPLEMENTAL DATA

Supplemental material can be found at <http://www.rnajournal.org>.

ACKNOWLEDGMENTS

We thank Mark Edson, Yiqun Zhang, and Dror Berel for usability testing of the software. This work was supported in part by a Program Project Development Grant from the Ovarian Cancer Research Fund, NIH grant P30 CA125123, and the Dan L. Duncan Cancer Center at Baylor College of Medicine.

Received May 19, 2008; accepted July 26, 2008.

REFERENCES

- Alvarez-Garcia, I. and Miska, E. 2005. MicroRNA functions in animal development and human disease. *Development* **132**: 4653–4662.
- Betel, D., Wilson, M., Gabow, A., Marks, D., and Sander, C. 2008. The microRNA.org resource: Targets and expression. *Nucleic Acids Res.* **36**: D149–D153.
- Cohen, A., Shmoish, M., Levi, L., Cheruti, U., Levavi-Sivan, B., and Lubzens, E. 2008. Alterations in micro-ribonucleic acid expression profiles reveal a novel pathway for estrogen regulation. *Endocrinology* **149**: 1687–1696.
- Creighton, C. 2008. Multiple oncogenic pathway signatures show coordinate expression patterns in human prostate tumors. *PLoS ONE* **3**: e1816. doi: 10.1371/journal.pone.0001816.
- Creighton, C., Kuick, R., Misek, D., Rickman, D., Brichory, F., Rouillard, J.-M., Omenn, G., and Hanash, S. 2003. Profiling of pathway-specific changes in gene expression following growth of human cancer cell lines transplanted into mice. *Genome Biol.* **4**: R46. doi: 10.1186/gb-2003-4-7-r46.
- Creighton, C.J., Cordero, K.E., Larios, J.M., Miller, R.S., Johnson, M.D., Chinnaiyan, A.M., Lippman, M.E., and Rae, J.M. 2006. Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors. *Genome Biol.* **7**: R28. doi: 10.1186/gb-2006-7-4-r28.
- Doniger, S., Salomonis, N., Dahlquist, K., Vranizan, K., Lawlor, S., and Conklin, B. 2003. MAPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* **4**: R7. doi: 10.1186/gb-2003-4-1-r7.
- Du, T. and Zamore, P. 2005. microPrimer: The biogenesis and function of microRNA. *Development* **132**: 4645–4652.
- Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**: D258–D261.
- John, B., Enright, A., Aravin, A., Tuschl, T., Sander, C., and Marks, D. 2004. Human MicroRNA targets. *PLoS Biol.* **2**: e363. doi: 10.1371/journal.pbio.0020363.
- Krek, A., Grün, D., Poy, M., Wolf, R., Rosenberg, L., Epstein, E., MacMenamin, P., da Piedade, I., Gunsalus, K., Stoffel, M., et al. 2005. Combinatorial microRNA target predictions. *Nat. Genet.* **37**: 495–500.
- Lewis, B., Shih, I., Jones-Rhoades, M., Bartel, D., and Burge, C. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.
- Lu, J., Getz, G., Miska, E., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B., Mak, R., Ferrando, A., et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**: 834–838.
- Ozen, M., Creighton, C., Ozdemir, M., and Ittmann, M. 2008. Widespread deregulation of microRNA expression in human prostate cancer. *Oncogene* **27**: 1788–1793.

- Rajewsky, N. 2006. microRNA target predictions in animals. *Nat. Genet.* **38 Suppl.:** s8–s13.
- Sethupathy, P., Megraw, M., and Hatzigeorgiou, A. 2006. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods* **3**: 881–886.
- Storey, J.D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**: 9440–9445.
- Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E., et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**: 15545–15550.
- Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci.* **99**: 6567–6572.
- Tusher, V., Tibshirani, R., and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.* **98**: 5116–5121.
- Volinia, S., Calin, G., Liu, C., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci.* **103**: 2257–2261.
- Xu, X., Zhao, Y., and Simon, R. 2008. Gene Set Expression Comparison kit for BRB-ArrayTools. *Bioinformatics* **24**: 137–139.